Tail Metrics For Network Performance Based on GPD and Mixture Modeling

John Lu

Nell Sedransk

National Institute of Standards and Technology

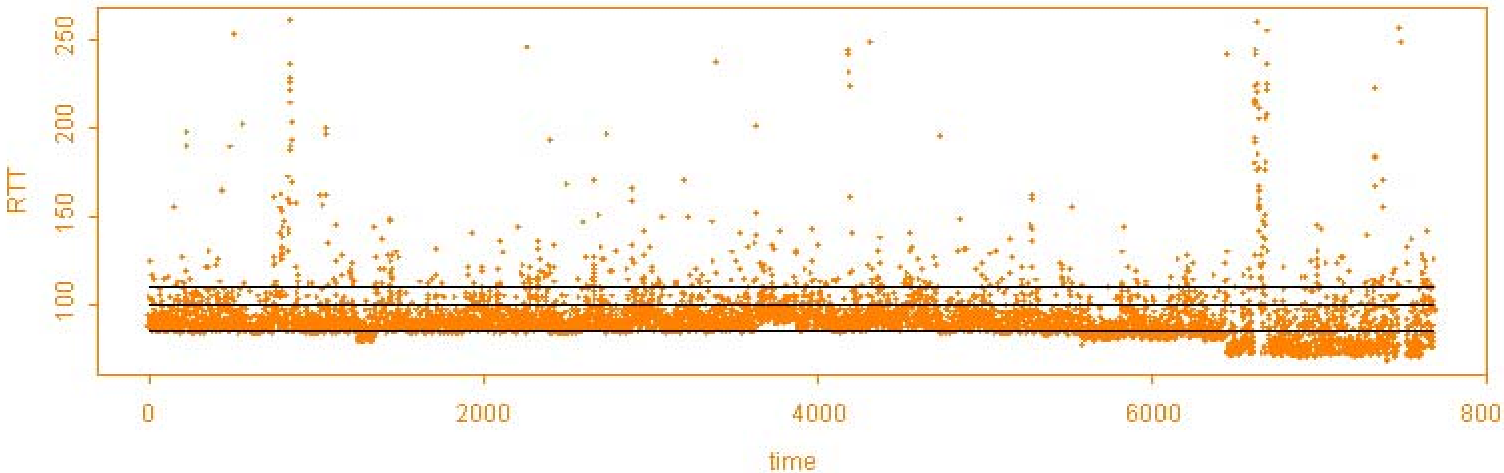
Part I: Motivation

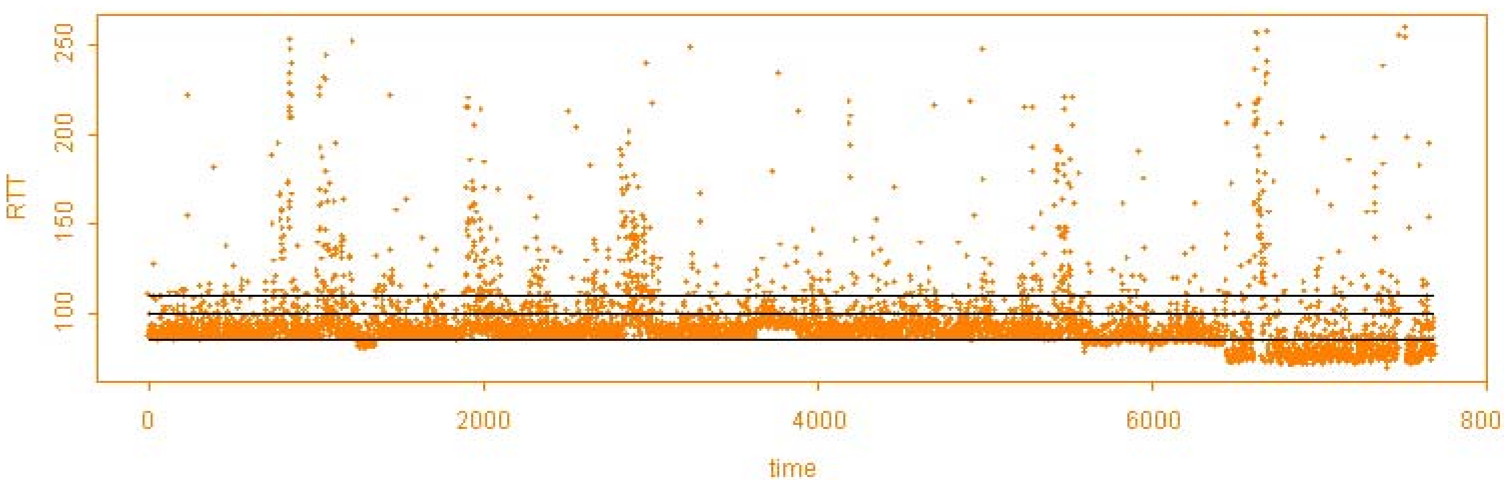
1. Network Performance metrics: latency (response delayness); jittering (variation), all related to the statistical characteristics of network data.
2. Heavy tailedness (power-law, self-similar in scale): myth or reality? Can QoS metrics be developed to take into account the tail distribution of data?
3. Does there really exist a multiscale approach to network traffic modeling in both scale and time?
4. Does there exist A statistical model which can apply to most network data, and in many different conditions?
5. Can statistical models be used to detect network changes or anomalies (e.g. DoS attacks)?

Part II: Modeling philosophy

1. Central statistics: standard statistics is about the center of distribution (e.g. mean, median), and the whole area of robust statistics is about avoiding effects from extreme values
2. Extreme statistics: extremes and tail values are of main interest in many problems, such as flooding, heat waves, risk analysis (finance, insurance, etc.), and as an example, we may define network slowness as RTT>100ms for certain packet size of certain dist.
3. Mixture distribution: The world is full of heterogeneity (not homogeneity), such as network change, time-varying, network attacks, etc.

RTT(measured at NIST and HP same time: Solid lines: RTT=85, 100, and 110 ms)





Traditional metrics fail!

	Before network change NIST	Before network change HP	After network change NIST	After network change HP
Median (Center)	89 ms	90 ms	80 ms	79 ms
Mean (affected by extremes)	94 ms	95 ms	108ms	98 ms
Mode (center)	86 ms	87 ms	73 ms	75 ms

Our proposed framework:

Mixtures with GPD

1. Mixtures: $f(x) = \sum_{i=1}^{m} p_i f_i(x)$ (density or CDF):

Example 1: piecewise parametric density fit at different scales: e.g. left tail to middle part by gamma or Weibull; right tail by Pareto or GPD; contamination, and to allow for tail perturbations

by mixing another density, such as log normal.

Example 2: kernel density estimation (extremely flexible): $f(x)= (1/n)\sum_{i=1}^{n} f(x_i - x)$, x_i data points.

smoothing of histogram--- flexible, data dependent, but has no physical explanation. (Provides a reference fit, but has no predictive ability.)

2. GPD: generalized Pareto:

CDF: $\mathbf{F(x)=1-(1-kx/a)^{1/k}}$, $\mathbf{(k\neq 0)}$;

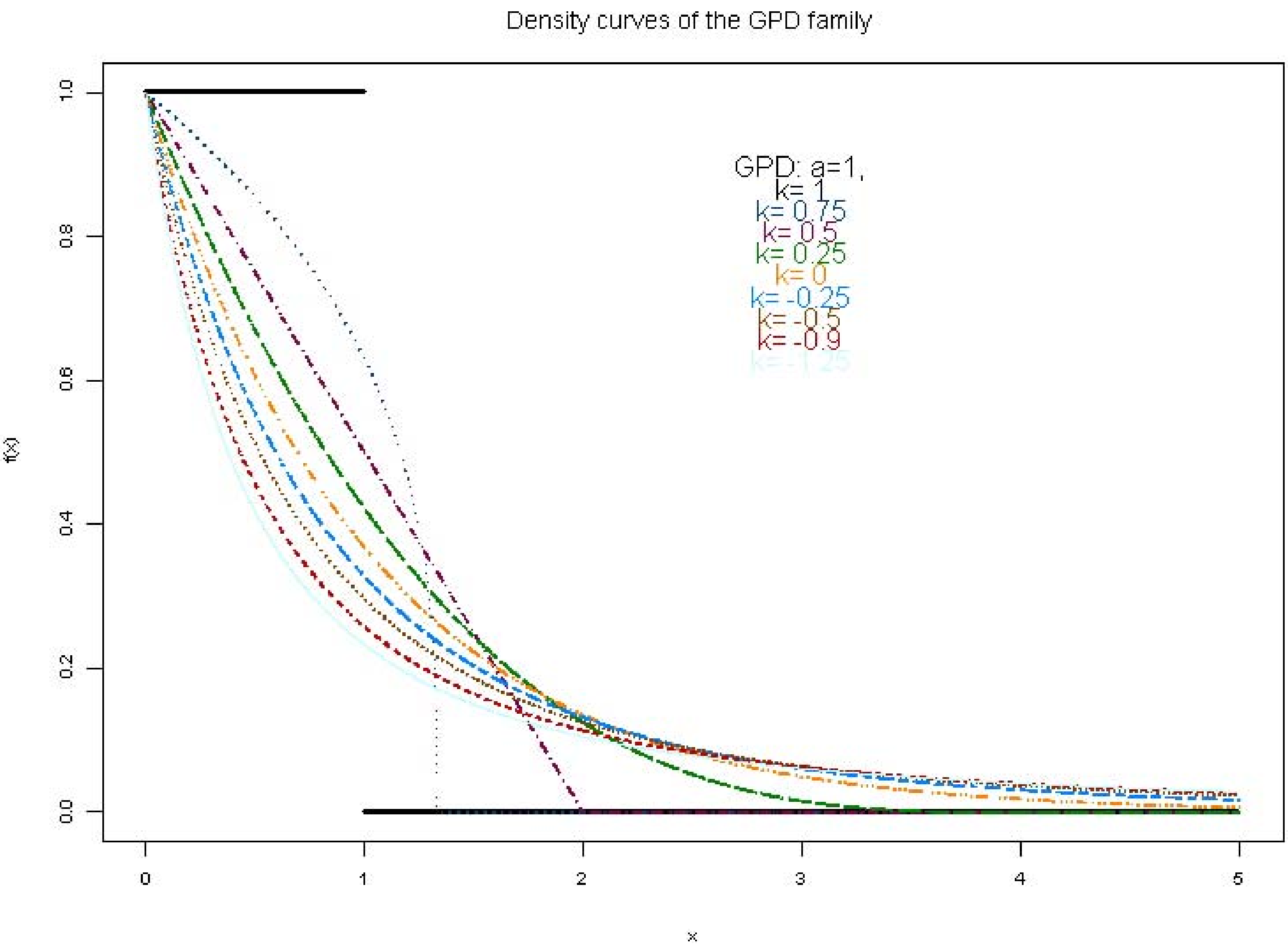
$\mathbf{1-exp(-x/a)}$, $\mathbf{k=0}$.

a: scale parameter, k: shape parameter

for the range of x, $0 \leq x < \infty$ for $k \leq 0$ and $0 \leq x < a/k$ for $k>0$. (Davison and Smith 1990)

- A simply and statistically justifiable model for large scale data: Based on sound probabilistic theory: apply very broadly to weakly dependent or even nonstationary process.
- threshold stability: stable at larger scale.
- Flexible enough to provide a model for tail distribution for almost all data, a test for powerlaw / exponential / heavytailness.

Rich behavior of GPD models



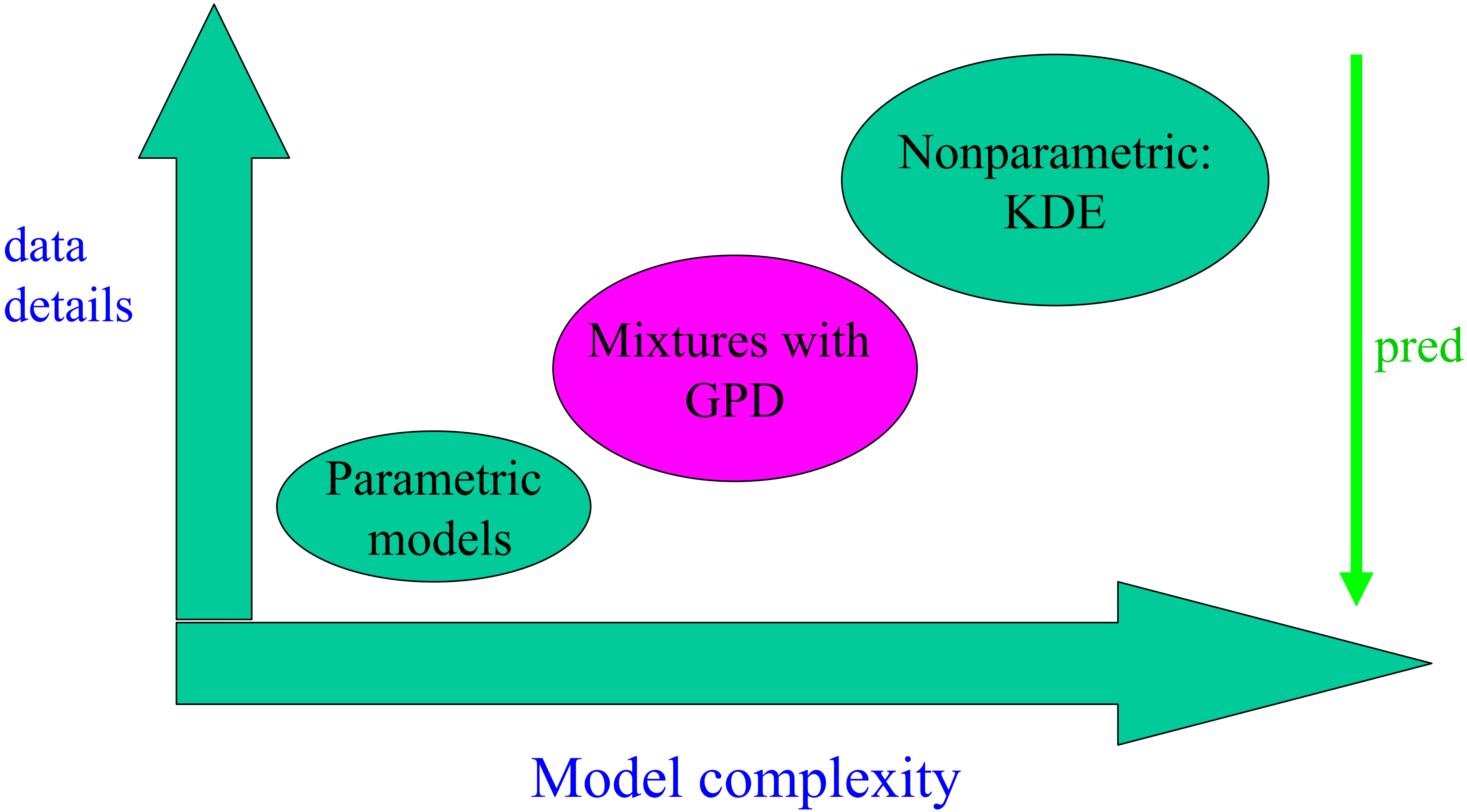
Why another model?

There are bewildering statistical distributions being fitted: log normal, Weibull, gamma, Pareto,……, why another?

Answers:

1. Existing models are ad hoc and you can always come up with a better model for a particular data set, but very hard to have a model which fits all data, or at least most network data under different conditions
2. There're simply not enough data to distinguish between Pareto and exponential in the tails.
3. Neither Pareto (power-law) distribution nor exponential distribution fits all parts of the data: both are monotone decreasing and will fail badly at the tails, if fitted to all data points.

Keep it simple: a model is as useful as data warrant it

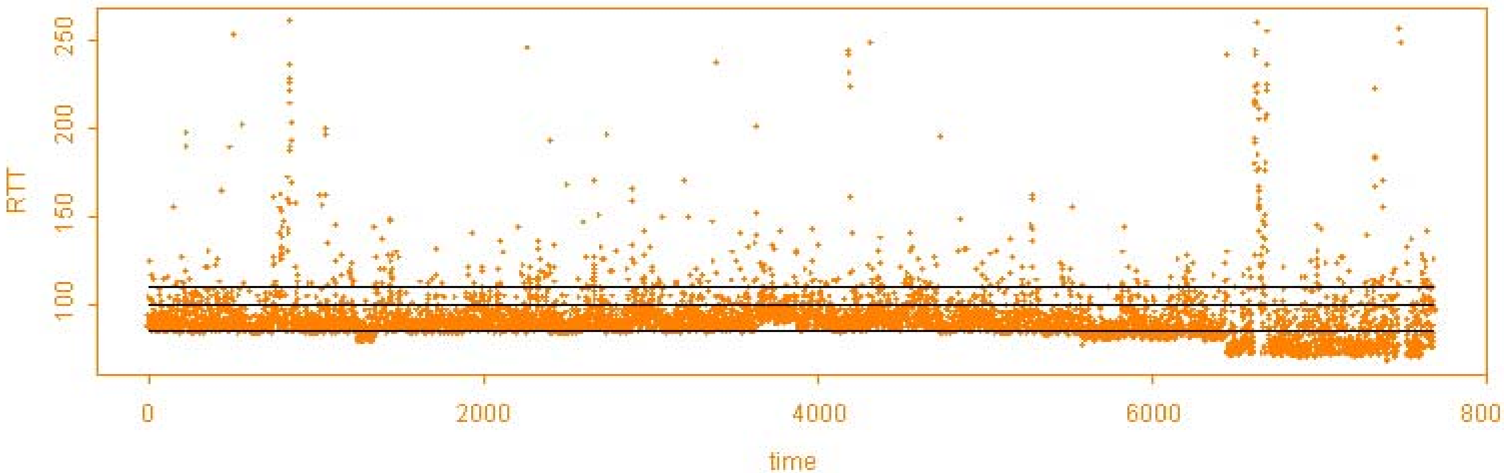


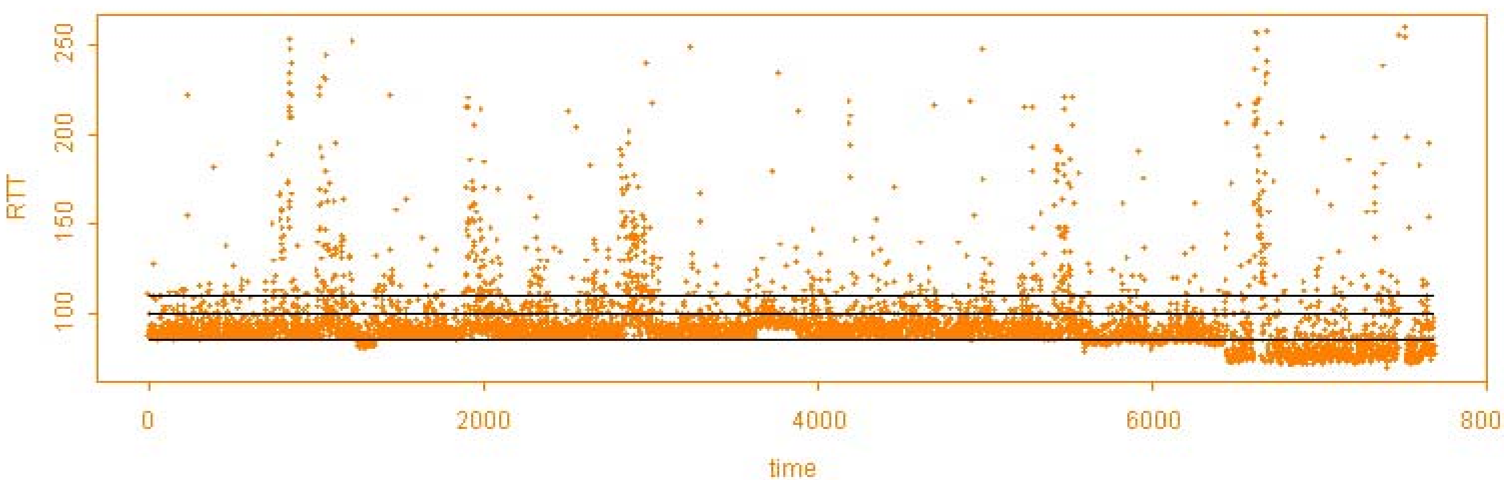
Part III: Tail data analysis

Example:

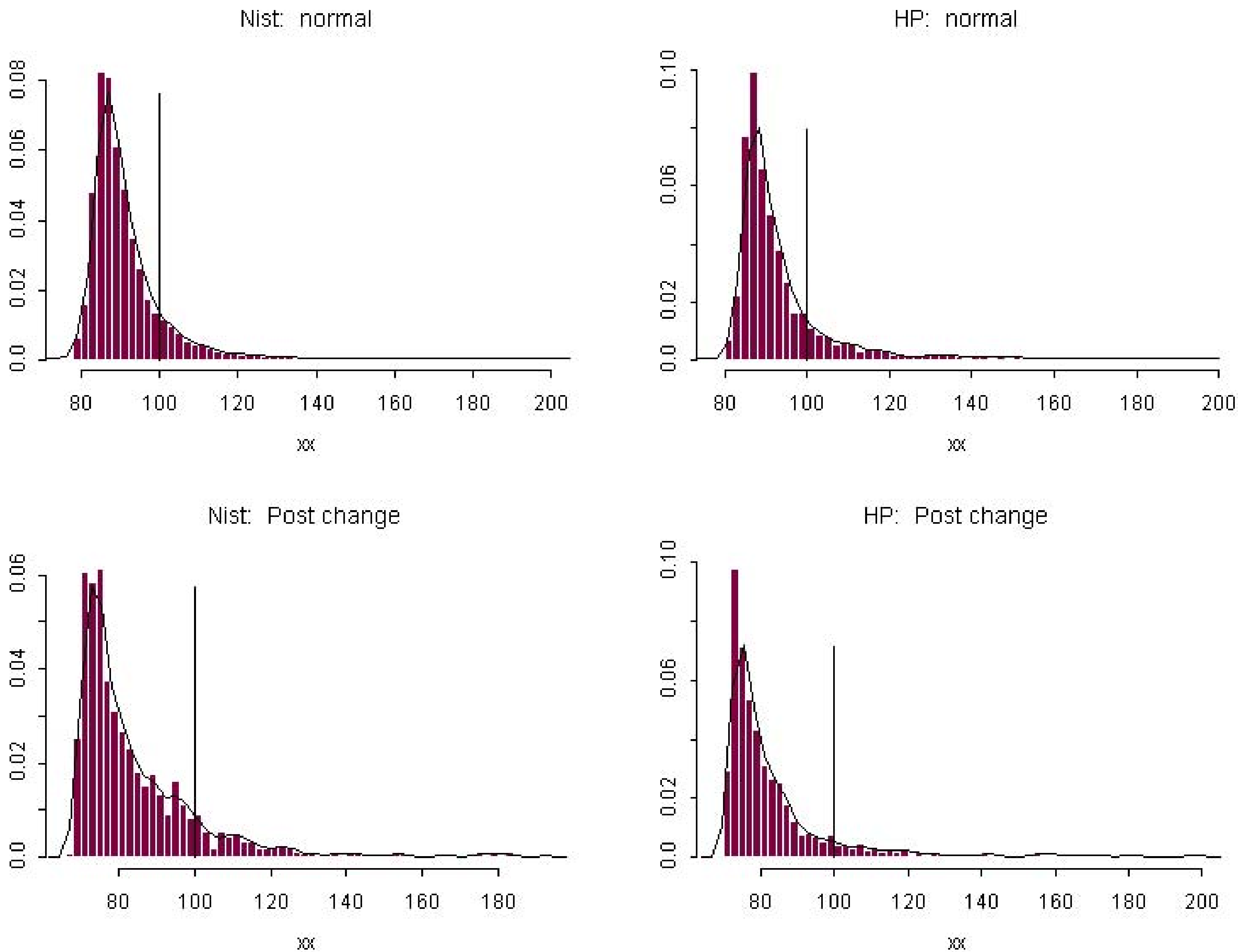
1. Round trip time (RTT, in ms, from PingER): NIST to HP, HP to NIST
2. Fairly long: 7692 points, contains long trend, heavy-tail, dependence, and
3. Network change around 6450; and this anomaly will be tested using our method.

RTT(measured at NIST and HP same time): Cases I, II, II: RTT>=85, 100, and 110 ms

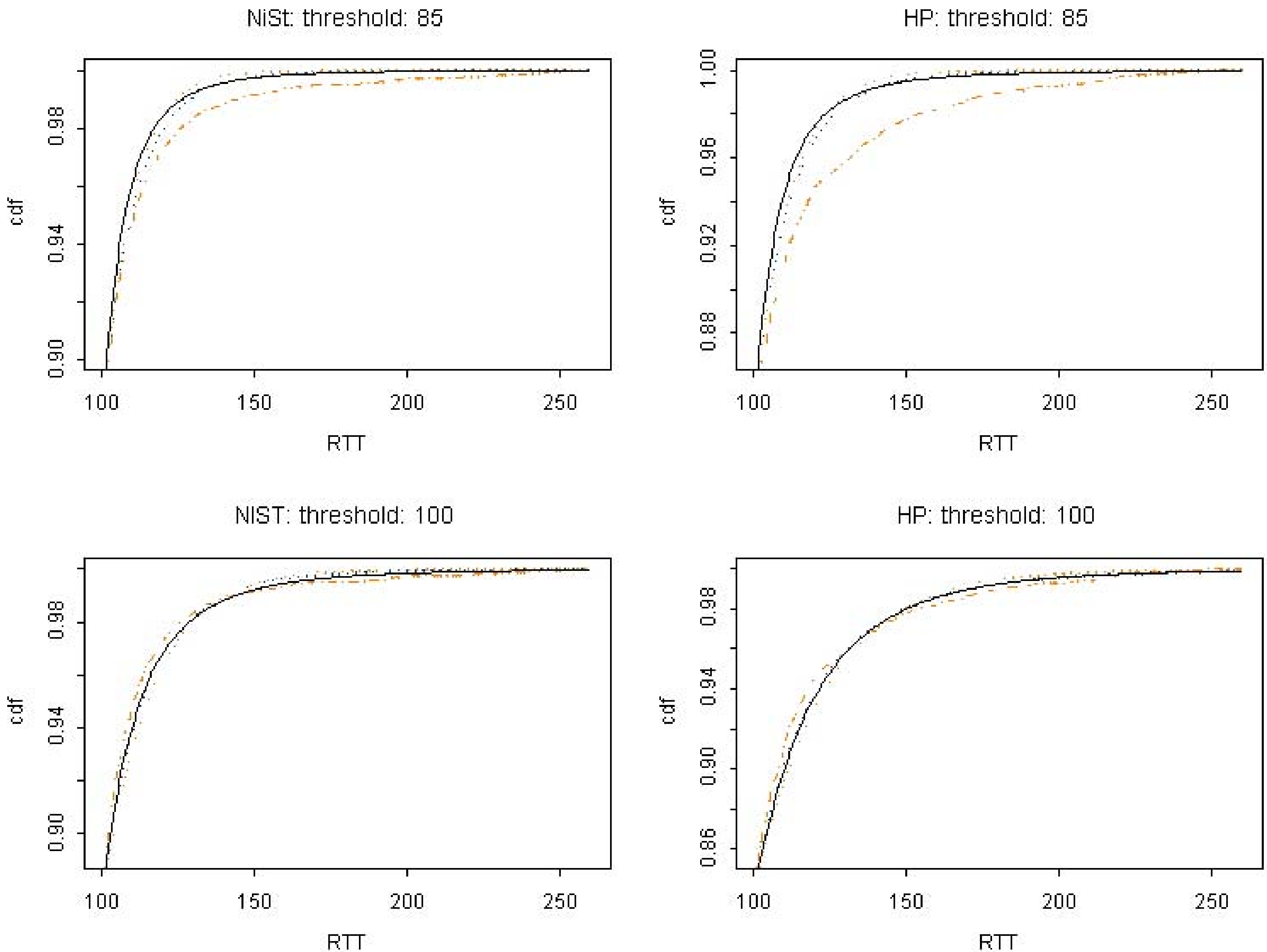




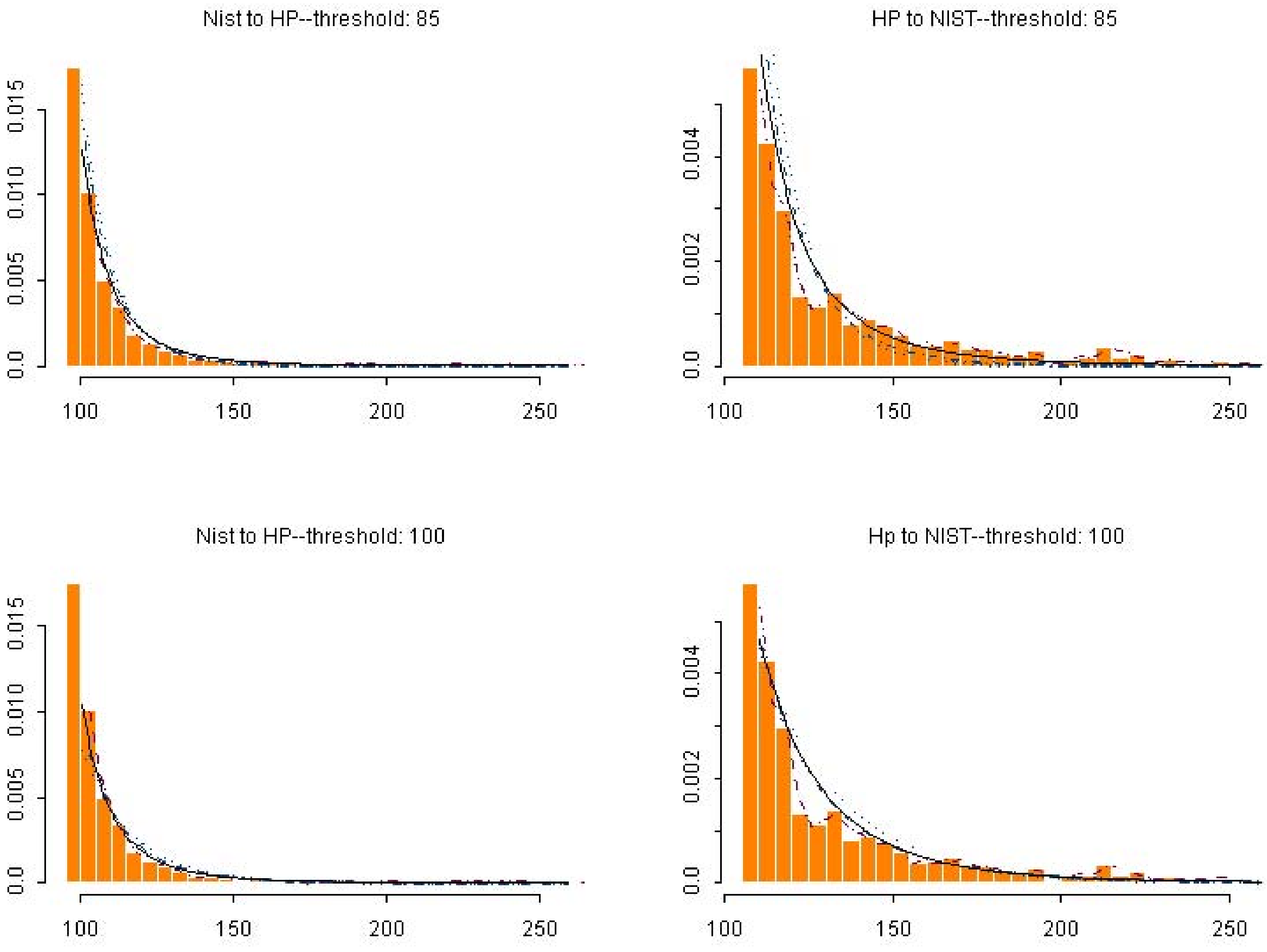
Histgram and density estimate of RTT



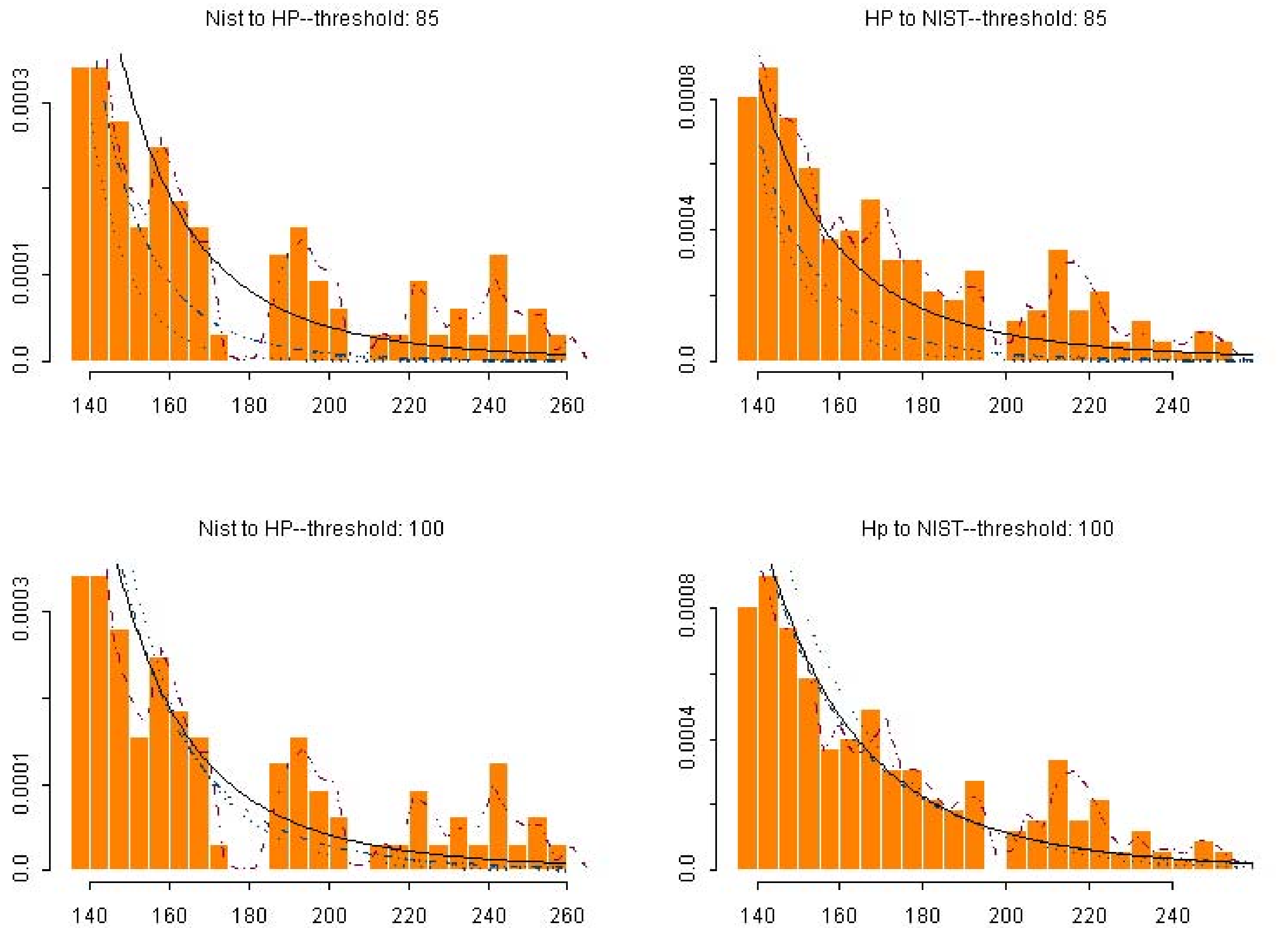
Tail modeling fit: in terms of CDF



Tail modeling: in terms of density



Tail density: closer look



Mid-point summary

- Any single family distribution (Pareto, exponential, GPD) fails to capture tails of data. GPD is doing better b/c it's richer.
- At tails, using only data above around the 90th percentile, all distributions (Pareto, exponential, and GPD) give very decent fit at tails.

  Heavy-tail ?: all power-law a (or 1/k) > 3, so not as heavy as generally thought.
- Above is for the normal network behavior. We note significant performance changes due to network change (at 6450). Let's look closely at the tails.

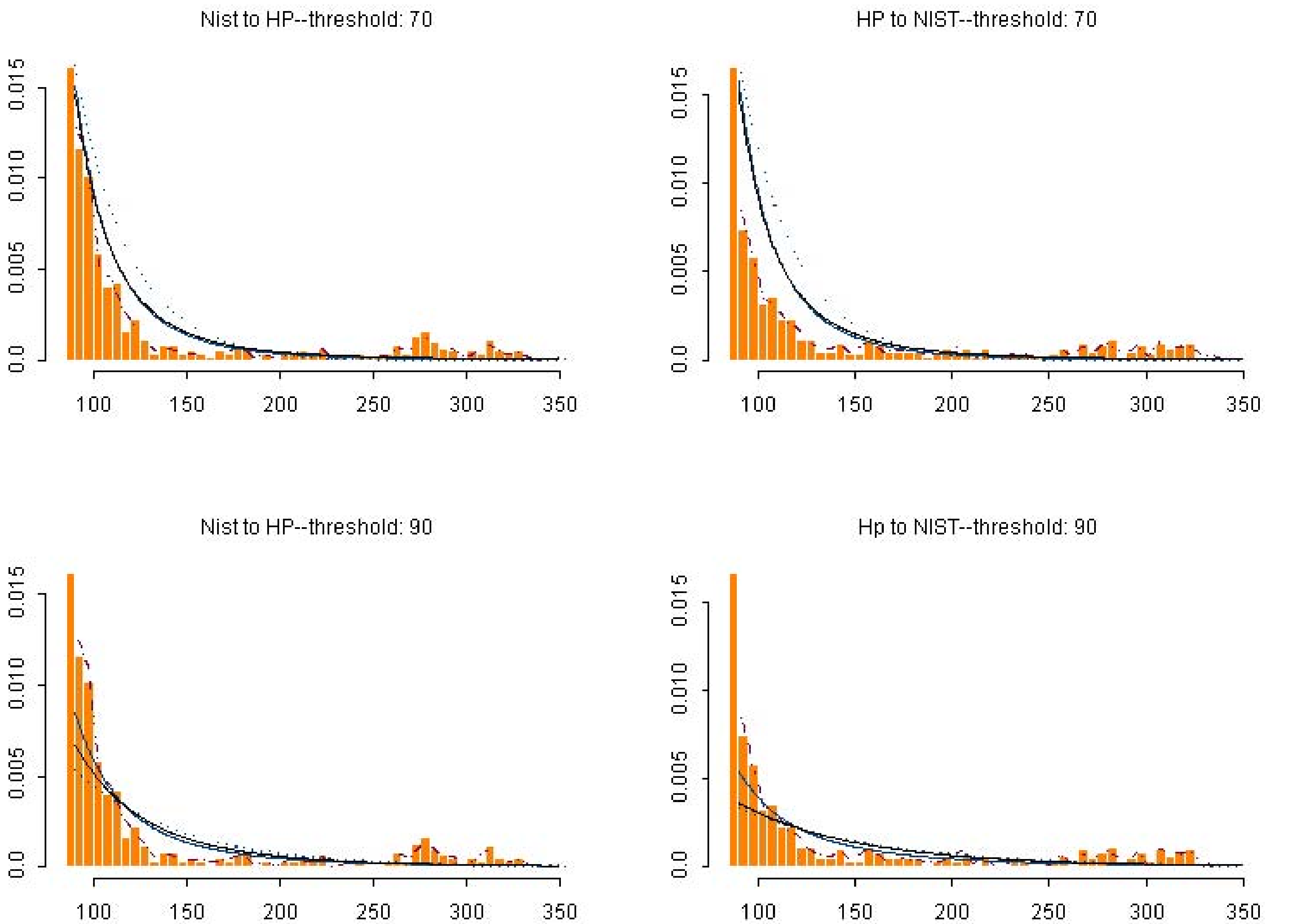
Part IV: Modeling tail contamination with mixtures

Using NIST-HP RTT data: post network change data (6450-7692) to illustrate

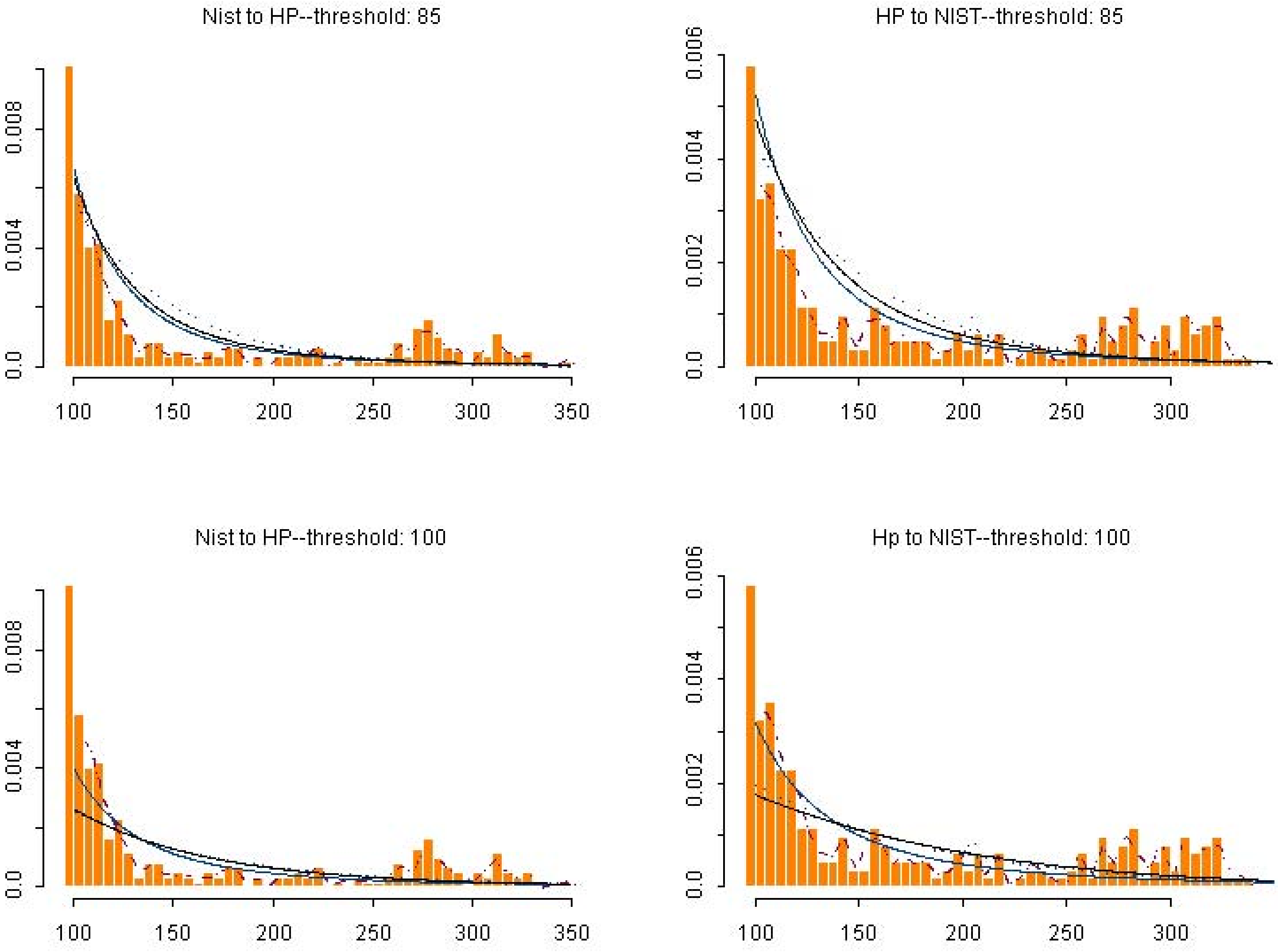
distribution of tails sensitive to network change.

1. Network faster but more volatile at tails
2. more frequency of network slowness, even if center of latency data may go down
3. Whether our model can detect external contamination and perturbation at tails?

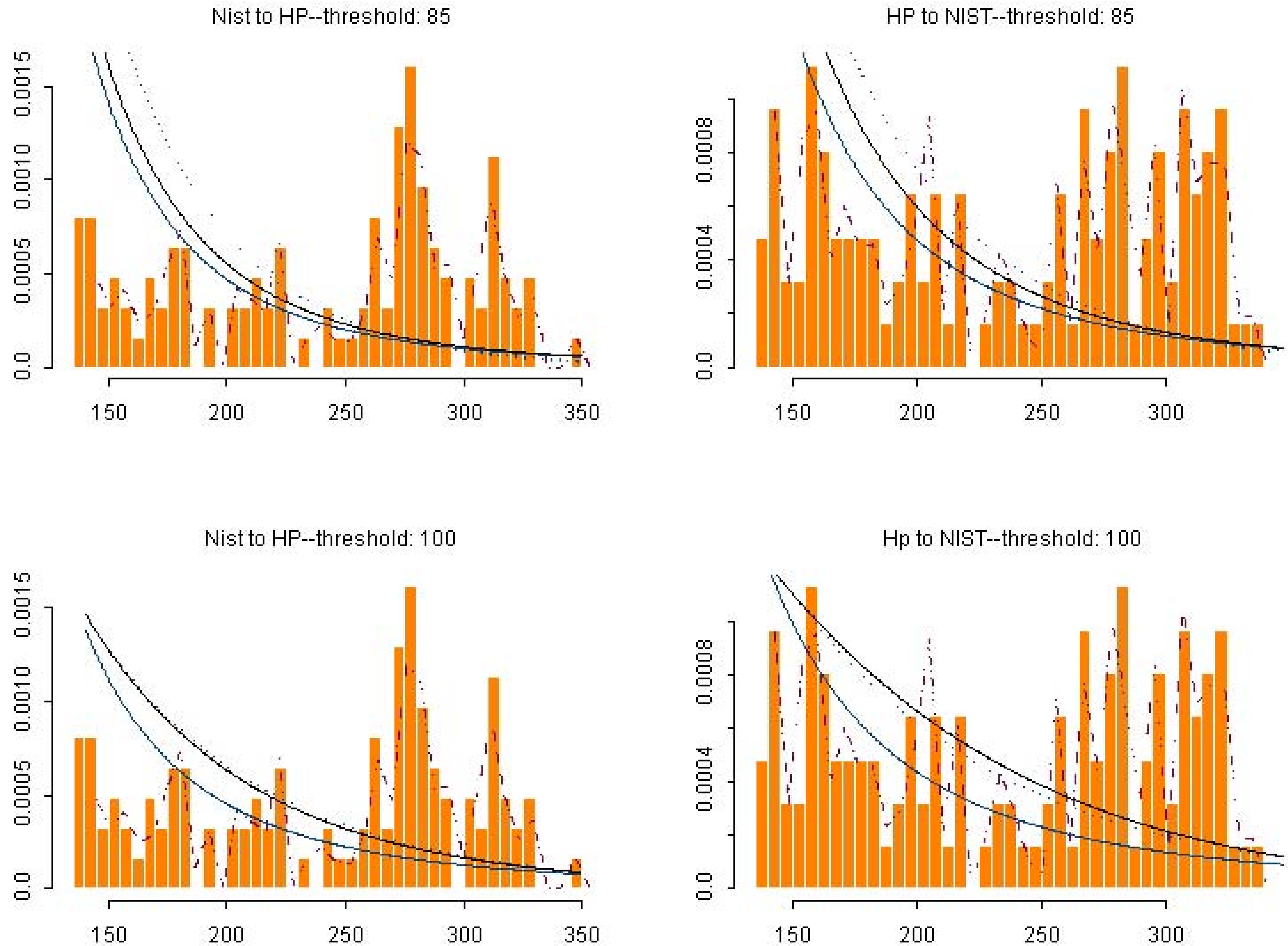
Tail density fit to post network change data (6451-7692):



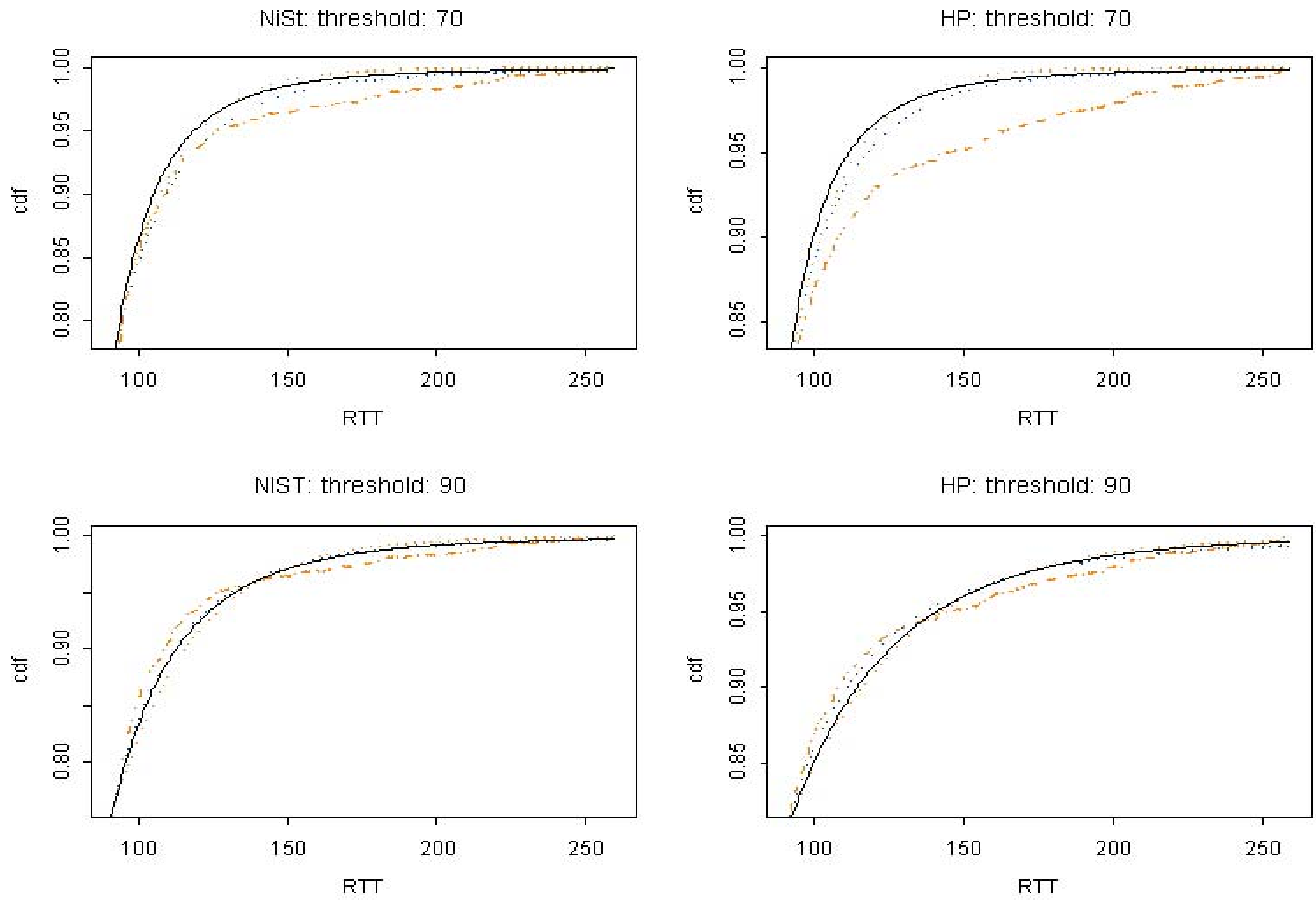
Tail density fit to post network change data (6451-7692) (continued)



(Cont.) A closer look at tail density



CDF tail fit for post network change data

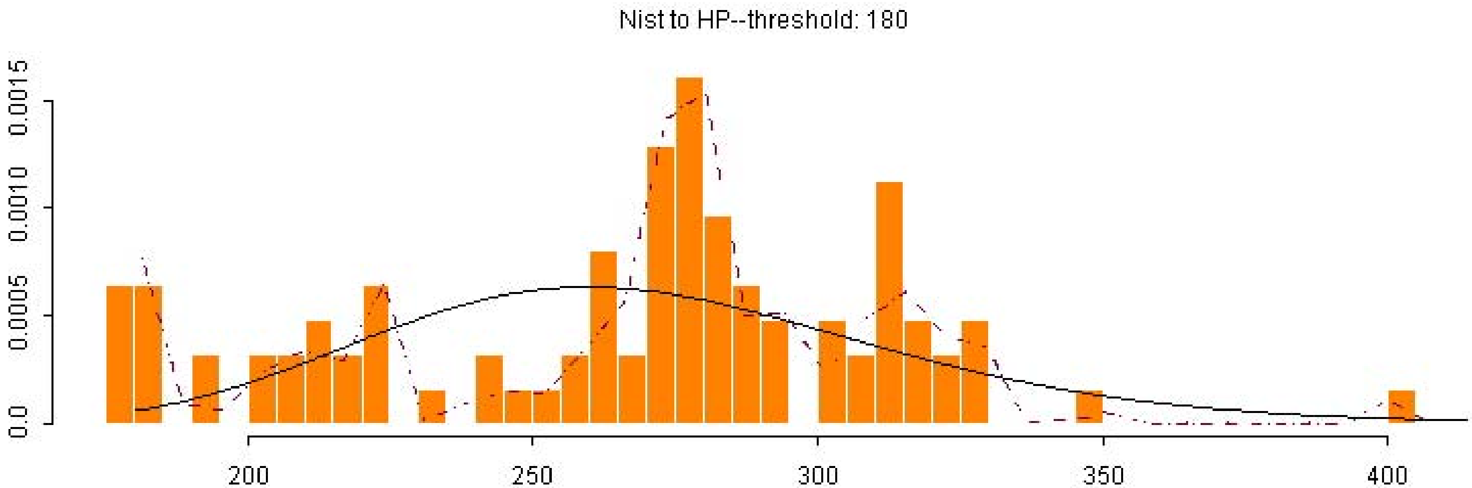


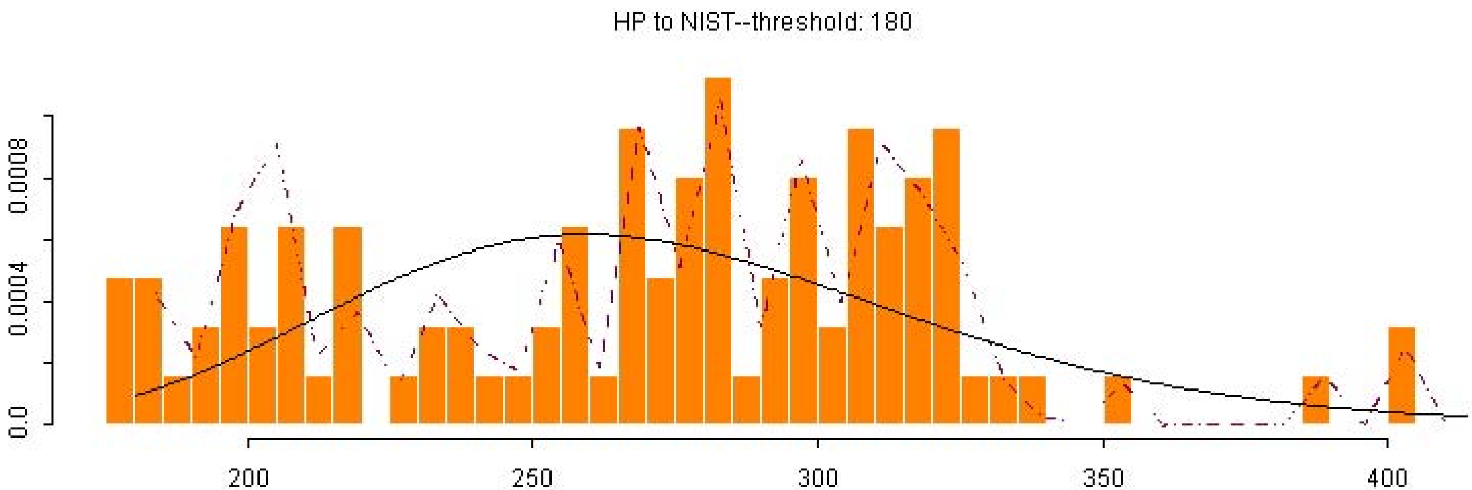
Comments

The fact that the GPD cannot model the tails, even if high threshold, indicate that there is an external contamination, perturbation in the tails.

(A consequence of threshold stability property of GPD.)

Model far tail contamination by adding a new mixing component: e.g. log normal

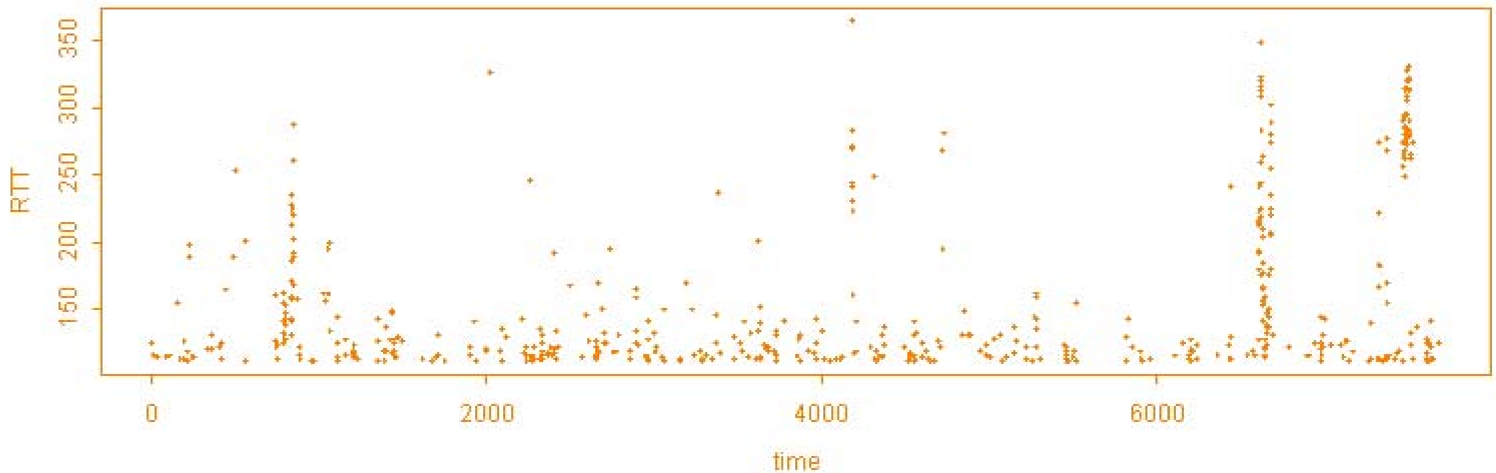


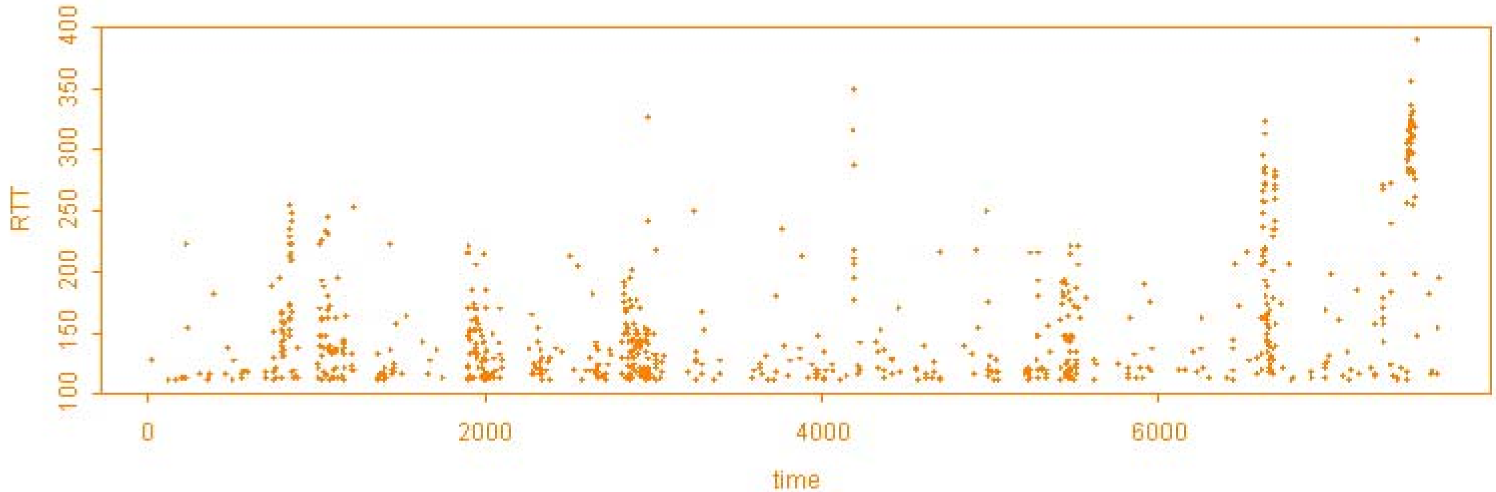


Comments:

1. Multiscale analysis in scale and time: zoom in at the tail behavior (extreme network slowness), and temporal network switching (non-stationarity)
2. A general tail model: There may not be enough data at the tails to distinguish between exponential and power-law.
3. We emphasize tail characteristics, which may not correlate with what’s going on in the main part of the data.
4. Contamination distribution at tails may come from a difference source and threshold stability of GPD will provide a detection at tails.

Take into account for temporal dependence:Excess delay RTT(>110ms)





Part V: Future work and potential applications

1. Apply to more data sets of different kinds
2. Further developments of GPD mixture modeling:
   - Temporal effects: clusters in exceedances
   - model explicitly network effects on tail metrics
3. QoS and network visualization using tail metrics:
4. Network traffic emulation (NIST NET): (Mark Carson):

   We work on enhancing statistical simulation models

5. Network security: early warning of network slowness

Conclusions

- Proposed a statistically and probabilistically justifiable approach for modeling tail distribution of network data
- Mixture framework allows for drastic changes in network behavior, in terms of model parameters, mixtures for contamination at tails

  The tail property of network measurements is most important and lends to novel tail performance metrics. For example, network's median performance improves, but tail behavior gets worse.

Comments and suggestions, papers exchanges, contact:

john.lu@nist.gov

nell.sedransk@nist.gov

Statistical Engineering Division, NIST

Gaithersburg, MD 20899-8980

David.Su@nist.gov

Advanced Network Technology Division, NIST